Structure and evolution of the gorilla and orangutan growth hormone loci

Article (Accepted Version)

Structure and evolution of the gorilla and orangutan *growth hormone* loci.

Antonio Ali Pérez-Maya[1], Michael Wallis[2], Hugo Alberto Barrera-Saldaña[1*].

[1]Departamento de Bioquímica y Medicina Molecular, Facultad de Medicina, Universidad Autónoma de Nuevo León, Monterrey, Nuevo León 64460, México.

[2]Biochemistry and Biomedicine Group, School of Life Sciences, University of Sussex, Falmer, Brighton, Sussex BN1 9QG, UK.

Keywords: Growth hormone, chorionic somatomammotropin, Gorilla, Pongo.

Short title: Gorilla & orangutan *GH* loci.

*** Correspondence:**

Dr. Hugo A. Barrera-Saldaña.

Departamento de Bioquímica y Medicina Molecular.

Facultad de Medicina de la Universidad Autónoma de Nuevo León.

Av. Madero y Dr. Aguirre Pequeño, Col. Mitras Centro.

Monterrey, Nuevo León, México 64460.

Tel.: +52 81 83294173 x2591; fax: +52 81 83337747.

E-mail: habarrera@gmail.com

ABSTRACT

In primates, the unigenic *growth hormone (GH)* locus of prosimians, expressed primarily in the anterior pituitary, evolved by gene duplications, independently in New World Monkeys (NWM) and Old World Monkeys (OWMs)/apes, to give complex clusters of genes expressed in the pituitary and placenta. In human and chimpanzee, the GH locus comprises five genes, GH-N being expressed as pituitary GH, whereas GH-V (placental GH) and CSHs (chorionic somatomammotropins) are expressed (in human and probably chimpanzee) in the placenta; the CSHs comprise CSH-A, CSH-B and the aberrant CSH-L (possibly a pseudogene) in human, and CSH-A1, CSH-A2 and CSH-B in chimpanzee. Here the *GH* locus in two additional great apes, gorilla (*Gorilla gorilla gorilla*) and orangutan (*Pongo abelii*), is shown to contain six and four *GH*-like genes respectively. The gorilla locus possesses six potentially expressed genes, *gGH-N*, *gGH-V* and four *gCSHs,* whereas the orangutan locus has just three functional genes, *oGH-N, oGH-V* and *oCSH-B*, plus a pseudogene, *oCSH-L*. Analysis of regulatory sequences, including promoter, enhancer and P-elements, shows significant variation; in particular the proximal *Pit-1* element of *GH-V* genes differs markedly from that of other genes in the cluster. Phylogenetic analysis shows that the initial gene duplication led to distinct *GH*-like and *CSH*-like genes, and that a second duplication provided separate *GH-N* and *GH-V*. However, evolution of the *CSH*-like genes remains unclear. Rapid adaptive evolution gave rise to the distinct *CSHs*, after the first duplication, and to *GH-V* after the second duplication. Analysis of transcriptomic databases derived from gorilla tissues establishes that the *gGH-N*, *gGH-V* and several *gCSH* genes are expressed, but the significance of the many *CSH* genes in gorilla remains unclear.

Introduction

Mammalian growth hormone (GH-N) is produced mainly in the anterior pituitary gland and stimulates the growth and metabolism of many tissues, including muscle, bone and cartilage, playing an important role in development. It is structurally related to a second pituitary hormone, prolactin (PRL) and, in primates, to the placental chorionic somatomammotropin (CSH; also known as placental lactogen) and variant GH (GH-V). During pregnancy, these placental hormones, produced primarily from fetal tissue, appear to act on maternal metabolism to secure nutrients and energy favoring the wellbeing of the fetus (Walker et al. 1991).

GH-N binds to a dimeric receptor located in the plasma membrane of its target cells. The hormone cross-links the two receptor molecules in the dimer, two distinct binding sites (sites 1 and 2) on GH-N binding to similar sites on the two receptor monomers. Determination of the 3-D structure of the dimeric extracellular domain of the human GH receptor bound to hGH-N allowed these binding sites to be defined in detail (de Vos et al. 1992).

The human *GH (hGH)* locus is localized on chromosome 17q24.2 and comprises five genes extending over 50 kb of DNA (Harper et al. 1982; Chen et al. 1989). From 5' to 3' the genes are: *hGH-N*, *hCSH-L*, *hCSH-A*, *hGH-V* and *hCSH-B*. *hCSH-A* and *hCSH-B* encode the same mature protein (Barrera-Saldaña et al. 1983). *hCSH-L* is substantially altered compared with the remaining *CSH* genes, including an altered splice site; it may be a pseudogene, though production of functional products has not been ruled out (Resendez-Perez et al. 1990; Misra-Press et al. 1994). *hGH-V* encodes a variant GH that appears to largely replace pituitary GH during late pregnancy. The divergent tissue-specific temporal patterns of expression of genes in

the *hGH* gene locus is controlled by diverse regulatory elements, including promoters and enhancers (Barrera-Saldaña, 1998). A distal locus control region (LCR) located -15 to -32 kb upstream of the locus also plays a crucial role (Su et al. 2000; Ho et al. 2004).

A complex *GH* locus is present in New World monkeys (NWM), Old World monkeys (OWM) and apes, but in most other mammals, including most prosimians (Adkins et al. 2001; Wallis et al. 2001), only a single *GH*-like gene has been found. The complex gene clusters found in higher primates presumably arose by several rounds of gene duplication, which occurred independently in NWM and OWM. In NWM gene numbers are variable and typically large, and they are separated by rather short, regular intergenic regions (IGRs) of about 2 kb (Gonzalez-Alvarez et al. 2006; Wallis and Wallis 2002, 2006). In OWM and apes the genes are separated by longer intergenic regions, of two types, respectively 6 kb and 13 kb in length (Chen et al. 1989; Gonzalez Alvarez et al. 2006; Pérez-Maya et al. 2012). In all cases the gene duplications resulted in multiple *GH*-like genes expressed in the placenta, the physiological role of which remains unclear.

The organization of the *GH* locus varies considerably among apes and OWM. Rhesus monkey and baboon have six genes/pseudogenes in the cluster (Gonzalez Alvarez et al. 2006; Rodríguez-Sánchez et al. 2010) while human and chimpanzee have five (Chen et al. 1989; Pérez-Maya et al. 2012). Genome sequences reported for other apes (Locke et al. 2011; Scally et al. 2012) suggest six *GH*-like genes in gorilla and four in orangutan, but in each case sequences reported for the *GH* gene cluster are incomplete, probably due to the repetitive nature of the locus. Here we re-examine data available from genomic studies, together with clones from BAC libraries, to

determine the complete sequences of the *GH* gene loci in gorilla and orangutan. We evaluate these data to throw further light on the nature of the *GH* clusters in these species and the evolution of *GH*-like genes in higher primates.

Material and methods

Amplification, molecular cloning and sequencing of the gorilla and orangutan *GH* loci.

A BAC clone (CH255-4K22; insert size about 233 Kbp) containing the gorilla *GH* (*gGH*) locus was obtained by screening a gorilla BAC library (CHORI-255; constructed from a male Western lowland gorilla "Frank"; average insert size 200 Kbp). A BAC clone (CH253-28b13; insert size about 203 Kbp) containing the orangutan *GH* (*oGH*) locus was obtained by screening an orangutan BAC library (CHORI-253; constructed from a male Sumatran orangutan "Segundo"; average insert size 177 Kbp). These BACs were obtained from the BACPAC Resource Centre (Children's Hospital Oakland Research Institute, Oakland, California; access generously granted by Dr Peter de Jong) by screening with a radiolabeled probe derived from the *hGH-N* gene. Extrachromosomal DNA from a bacterial stock carrying this BAC was extracted with the BACMAX DNA Purification Kit, following the manufacturer's instructions (EPICENTRE, Madison, WI, USA). Screening, PCR primer design, PCR amplification conditions, and cloning and sequencing of amplicons were carried out as previously described (Pérez-Maya et al. 2012). Sequencing of the BACs was performed by McGill University and the Génome Québec Innovation Centre, Montreal, Canada, using a Roche Genome Sequencer FLX 454. These clones proved, in each case, to contain the complete *GH* locus plus flanking genes.

Genomic assembly

The final assembly of the genomic sequences for the gorilla and orangutan *GH* loci was achieved using the approach previously described (Wallis 2008; Pérez-Maya et al. 2012). Using BLAST and BLAT search methods (Altschul et al. 1990; Kent 2002), human *GH* gene regions were used as query sequences to identify sequences from the WGS trace archive (http://www.ncbi.nlm.nih.gov/Traces). Trace sequences were generated at Sanger Centre and NIH intramural sequencing center (gorilla) and Baylor College of Medicine and Washington University Genome Sequencing Center (orangutan). These were then integrated with the sequences derived in this study. Sequences from trace files and 454 reads were assembled using CodonCode and the Staden Package (https://sourceforge.net/projects/staden/). The assembled sequences were evaluated by comparing with published sequences of the chimpanzee (JN622009) and human (J03071) *GH* loci. The assembled gorilla and orangutan *GH* loci sequences have been deposited in GenBank/EMBL/DDBJ databases, with accession numbers KT971340 and KT959234.

Sequence analysis

The nucleotide sequences obtained in this work were aligned using the ClustalW method (Higgins and Sharp 1988) followed by manual adjustment where necessary. Protein sequences were derived by conceptual translation of the coding sequences. Other *GH* gene sequences retrieved from GenBank were also used in the analyses. Phylogenetic analysis was carried out

using parsimony, neighbor-joining, and maximum-likelihood methods in the programs MEGA5 (Tamura et al. 2011) and PAUP* (Swofford 1998). Evolutionary rates for nonsynonymous (dN) and synonymous sites (dS) in coding sequences were determined using codeml in the paml package of Yang (2007), with a defined phylogenetic tree. The significance of amino acid differences was evaluated by molecular modeling using the hGH:receptor model of de Vos et al. (1992) and Rasmol or Pymol (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC). The GENECONV program (Sawyer 1999) was used to investigate gene conversion. GENECONV was run with mismatches allowed (setting g1 or g2). *p* values were corrected for multiple hypothesis testing.

Expression in transcriptomic libraries

Transcriptomic libraries from various gorilla and orangutan tissues in the sra archive (http://www.ncbi.nlm.nih.gov/Traces/sra) were screened online by BLAST analysis (blastn; expect threshold 10; word size 11; match/mismatch scores 2, -3; gap costs existence:5, extension:2; low complexity regions filtered) using the *hGH-N* coding sequence (CDS) as Query. The identity of hits was determined by comparing with available sequences. Useful results were obtained from projects SRP007412 (Brawand et al. 2011; gorilla and orangutan), SRP017959 (Necsulea et al. 2014; gorilla) and SRP021223 (Pipes et al. 2013; gorilla).

Results and Discussion

Organization of the gorilla and orangutan *GH* loci.

The repetitive nature of the *GH* locus in primates, resulting from several rounds of gene duplication, makes it difficult to achieve accurate genomic assembly. This was seen in our previous study on the locus in chimpanzee containing five *GH*-like genes (Perez-Maya et al. 2012). Publications on the genomes of gorilla and orangutan (Locke et al. 2011; Scally et al. 2012) indicate six and four *GH*-like genes respectively, but incomplete regions remain within the sequences for these loci. In order to facilitate detailed analysis of the evolution of the *GH* locus in higher primates we have re-examined the sequences, using both data available from the WGS databases and from BACs containing the entire gorilla and orangutan *GH* gene clusters.

Combining these approaches we obtained complete sequences for the gorilla and orangutan *GH* gene clusters (see Supplementary Figs 1 and 2 where the gene sequences are marked up). These completed sequences show that, as previously proposed, the gorilla *GH* locus contains six genes and the orangutan *GH* locus contains four genes, as summarized in Fig. 1. Overall similarity between the six genes of gorilla is about 94% and between the four of orangutan about 91%, reflecting their recent generation by multiple rounds of gene duplication (Chen et al. 1989). As in human and chimpanzee, the genes flanking the loci are *CD79b* and *TCAM*. These hominid *GH* gene clusters differ in detail in a number of respects, as discussed below.

Encoded proteins

Pituitary GHs

Mature proteins (i.e. excluding signal peptide) encoded by the gorilla and orangutan *GH-N* genes are identical to human GH-N (Fig. 2). They differ from chimpanzee GH at one residue (residue 126 is Glu in chimpanzee, Gly in human, gorilla, and orangutan). More differences are seen in the signal peptides where residue -24 is Thr in human, Pro in chimpanzee, Ala in gorilla and orangutan and residue -13 is Gly in human, chimpanzee and gorilla, Ala in orangutan.

Placental GHs

The gorilla *GH-V* gene encodes a mature protein that differs from human and chimpanzee GH-V at three positions (Fig. 2): Ile-3, Asp-21, and Ser-146 in gorilla replace Thr-3, Tyr-21, and Phe-146 in both chimpanzee and human. Residue 21 (His in hGH) is within binding site 1 in the hGH:receptor structure of de Vos et al. (1992), while residue 3 (Thr in hGH) is close to binding site 2. The orangutan *GH-V* gene encodes a mature protein that differs from hGH-V at 11 residues (see Fig. 2). In some respects it is more similar to GH-N than are other GH-Vs, possibly reflecting gene conversion (see below).

Placental CSHs

The *CSH*-like genes of gorilla and orangutan are more variable than the *GH* genes, but the most striking variation is in the number of such genes, varying from two in orangutan to four in gorilla. Human *CSH-L* is a potential pseudogene (see Introduction), having an altered 5' splice donor site at intron 2, which substantially affects the derived mRNA (Resendez-Pérez et al. 1990). *CSH-L* in orangutan, following *GH-N,* also appears to have an altered 5' splice donor site at intron 2. The likelihood that orangutan *CSH-L* is effectively a pseudogene is increased by an

in-frame stop codon in exon 5. *CSH-B* of orangutan has the hallmarks of a functioning gene; its sequence is clearly related to *hCSHA/B*, though the proteins encoded differ at 12 residues.

The altered splice site seen in *hCSH-L* is not seen in the equivalent genes in gorilla and chimpanzee, and in gorilla all 4 *CSH* genes are potentially translatable to give similar but distinct proteins. Nevertheless, the sequence of gorilla CSH-A1 does differ substantially from that of gCSH-A2, gCSH-A3 or gCSH-B (at about 10 residues) and in some respects resembles that of orangutan CSH-L. gCSH-A2, gCSH-A3 and gCSH-B are very similar, differing by only one or two residues, which are not located close to a receptor-binding site in the hGH-receptor model. The similarity between these three CSHs in gorilla may in part reflect gene conversion events as well as recent gene duplications in the *GH* locus (see below).

Regulatory elements

Several key regulatory elements have been identified that influence expression of the genes composing the human *GH* locus, including the proximal promoters, enhancers downstream of the *CSH* genes, upstream "P-elements" for both *GH-V* and *CSH* genes, and a locus control region far upstream of the *GH-N* gene (Ho et al. 2004). All but the last of these, which falls outside the limits of the *GH* locus considered here, are considered in detail below.

Proximal promoter

Various response elements are found in the promoter region of *hGH-N* (Eberhardt et al. 1996). In Fig. 3 they are compared with the corresponding regions of the other genes of the human *GH* locus and with the genes of the *cGH, gGH,* and *oGH* loci. The TATA box and the *Sp-1* element (Lemaigre et al. 1989) are completely conserved in all genes. The initiator binding site (*InrE*), which is required for efficient activity of the promoter and maximum activity of the enhancer (Jiang et al. 1995), is generally well conserved, as is the cyclic AMP response element (*CRE*, Eberhardt et al. 1996), contrasting with the situation in non-primate mammals (Wallis et al. 2001).

The sequences of both distal and proximal *Pit-1* elements are generally strongly conserved, with those of gorilla *gCSH-A2*, *gCSH-A3*, and *gCSH-B* genes, orangutan *oCSH-B* gene and chimpanzee *cCSH-A1* gene all being identical to that of the *GH-N* gene. A striking exception to this conservation of *Pit-1* elements is seen in the proximal element for the *GH-V* genes, which shows five nucleotide substitutions, shared in orangutan, gorilla, chimpanzee, and human. This confirms the previous conclusion (Perez-Maya et al. 2012) that this element underwent a period of rapid adaptive evolution after it arose by duplication of *GH-N*. This presumably reflects the low-level expression of GH-V in the placenta, compared with high level expression of GH-N in the anterior pituitary.

The thyroid hormone response-element (*TRE*; Glass et al. 1987) appears to be important in placenta rather than pituitary (Leidig et al. 1992). Possibly reflecting this, the *TRE* before the pituitary-expressed *GH-N* gene shows considerable sequence variation while that before *GH-V* is completely conserved. The *TREs* preceding *CSH* genes are fairly variable, though for most of

them positions -116, -112 and -110 are G, G and A, respectively; these sites were identified as crucial for the functioning of this element in *hCSH-A/B* (Leidig et al. 1992). In the case of chimpanzee, gene conversion events may have influenced variation in this region (Perez-Maya et al. 2012), and the same may be true for gorilla (see below). The variability seen in the *TRE* suggests that the pattern of responsiveness to thyroid hormones for these placental hormones may vary considerably.

The placental enhancer

An alignment of the enhancer that is found ~2.2 kb downstream of *CSH* genes is shown in Fig. 4, emphasizing the four regions identified by Jacquemin et al. (1994) by footprinting, of which regions *DF3* and *DF4* appear to be the most important. It has been demonstrated that insertion of an AGAA sequence upstream of the enhancer's DF-2 domain decreases its functionality (Jacquemin et al. 1994). Interestingly, this 4-base pair insertion is not present in the enhancer copies following the *hCSH-B*, *cCSH-B*, *gCSH-A3*, and *gCSH-B* genes. The enhancers of the remaining *CSH* genes show this insertion suggesting their reduced functionality (Jacquemin et al. 1994).

Besides this insertion there are a few substitutions that may affect these gene enhancers (Lytras et al. 1996; Jacquemin et al. 1996) and these are identified in Fig. 4. A substitution in position 208 (G for A), in the DF-3 domain, is found in all *CSH* genes of orangutan, gorilla and chimpanzee, suggesting that the enhancer here is much less active than in human *CSH-B*. Another substitution at position 228, T for A, may lower the activity of the enhancer in all *CSH* genes, except *hCSH-B*, *cCSH-B*, *gCSH-B*, *oCSH-L* and *oCSH-B*.

An interesting feature of the enhancer that does not appear to have been noted previously was revealed by analysis of the *GH* locus using the programme RepeatMasker (Smit et al. 1996–2010). This showed that each of the enhancer elements following the three *CSH* genes overlapped a *MER5B* transposable element (TE), a member of the hat-Charlie superfamily of DNA transposons. As indicated in Fig. 4, the overlap includes the DF3 and DF4 domains, which are the main contributors to enhancer activity (Jacquemin et al. 1996). A conserved *MER5B* element is also found following the *GH* gene in prosimians and dog, where there is a single gene in the *GH* locus (see Introduction). This suggests that the element was present long before the duplications of the *GH* gene that gave the *GH* gene cluster in higher primates and was subsequently co-opted (or exapted) to play a role in regulating *CSH* expression. It has been proposed previously that exaption of TEs to provide regulators of gene expression played a major role in the evolution of the maternal side of the placenta (Emera and Wagner 2012; Lynch et al. 2011, 2015). The contribution of *MER5B* to the *CSH* placental enhancer suggests that the same may apply for the fetal side, though the extent of this has yet to be determined.

P-element

The "P-element" is found upstream of genes of the human *GH* locus that are expressed in the placenta, but not *GH-N*. It is thought to inhibit expression of these genes in the pituitary (Norquay et al. 2006) and/or activate their expression in the placenta (Elefant et al. 2000; Ho et al. 2004). Equivalent sequences are found in the gorilla and orangutan, and may well serve the same function. It is noteworthy, however, that a P-element-like sequence is found upstream of

the *GH-N* gene in many other mammals, including marmoset and dog (Wallis and Wallis 2006), making its role as a pituitary suppressor less clear-cut.

An alignment of the region including the P-element is given in Fig. 5, which shows the three functional domains identified by Norquay et al. (2003, 2006). Domains PSE-A and PSE-B are highly conserved, while the PSE-C domain shows a number of substitutions largely shared between human, chimpanzee, gorilla and orangutan. Putative binding sites for HNF-3 and C/EBP (Norquay et al. 2006) are also largely conserved.

Glucocorticoid response element

A glucocorticoid response element (GRE) has been reported in the first intron of *hGH-N* and *hGH-V* (Slater et al. 1985). It is completely conserved in chimpanzee *GH-N* (Perez-Maya et al. 2012) but not *GH-N* of gorilla or orangutan or any of the *CSHs*. Interestingly it is fully conserved in the *GH-V* genes of all four hominids, suggesting that it may play a role in regulation of this placental gene.

Gene conversion

Gene conversion events played a role in the evolution of both the human and chimpanzee *GH* gene clusters (Chen et al. 1989; Perez-Maya et al. 2012). The GENECONV program (Sawyer 1999) was used to explore the occurrence of gene conversion events during evolution of the cluster in orangutan and gorilla, using in each case alignments of ~2700 nucleotides (nt) long comprising the *GH/CSH* genes from ~930 nt upstream of the start codon to ~220 nt downstream of the stop codon. Results are shown in Table 1.

In the case of gorilla, extensive gene conversions between *CSH-A1* and *CSH-A2*, *CSH-A3* and *CSH-B* are indicated. Statistical support is strongest for the *CSH-A1*:*CSH-A2* conversion, and the apparent *CSH-A1*:*CSH-A3* and *CSH-A1*:*CSH-B* conversions may reflect the similarity between *CSH-A2*, *CSH-A3* and *CSH-B* rather than separate conversion events. A conversion between *GH-N* and *CSH-A1* in the 5' utr, including the proximal promoter region, is also suggested, but statistical support for this is rather weak and the alignment of proximal promoters (Fig. 3) does not support it.

For orangutan, one or two gene conversions between *GH-N* and *GH-V* between exon 2 and intron 3 are indicated (Table 1). Statistical support is rather weak, but such conversion would accord with the similarity between the protein sequences in this region (Fig. 2).

The length of the alignment used for these gene conversion studies was limited because at greater lengths sequence similarity between *GH*s and *CSHs* is lost. A much longer alignment is possible if just *CSHs* are considered, and gene conversion was therefore investigated for an alignment of *CSH* sequences for all four hominids, comprising ~16090 nt extending from ~4100 nt upstream from the initiating ATG to ~10400 nt downstream of the stop codon. Results are shown in Table 2. For gorilla long sequences including all or part of the coding sequences were again identified as undergoing gene conversions, with some shorter conversions far upstream or downstream. For orangutan only a single gene conversion event was identified, well downstream of the CDS.

Gene expression

In human, genes of the *GH* locus are expressed mainly in anterior pituitary and placenta, but also to a lesser extent in some peripheral tissues, especially testis and ovary (Pérez-Ibave et al. 2014). Limited information is available about expression of GH-related genes in hominids, apart from human, but some transcriptome libraries are available, and potentially useful. Gorilla transcriptomic libraries available in the sra database (http://www.ncbi.nlm.nih.gov/Traces/sra) were screened by Blasting with *hGH-N* CDS, and hits obtained were ascribed where possible to members of the *GH* locus. Results are shown in Table 3. Transcriptomes giving a modest number of hits derived from testis (two separate projects) and a mixture of tissues, including ovary but not pituitary, placenta or testis. In testis very low expression levels for *gGH-N* and *gGH-V* were detected, but levels were higher for *CSHs*. In most cases hits could not be ascribed to individual *CSHs*, but in both testis databases most hits were to *CSH-A2*/*CSH-A3*/*CSH-B* and a few could be ascribed specifically to *CSH-B*; only a single hit that could be clearly ascribed to *CSH-A1* was detected. In the transcriptome derived from mixed tissues, a fair number of hits could be ascribed to *gGH-N* and *gGH-V*, but few to *gCSHs*. The results provide a preliminary demonstration that many of the six genes found in the gorilla *GH* cluster can be expressed, but further studies will be required to explore relative expression levels and establish whether all four potential CSH proteins are in fact produced. Blast screening of available orangutan databases detected no more than one hit for any transcriptome - too few to justify further consideration.

Evolution of the *GH* locus in great apes

Gene duplication

Chen et al. (1989), on the basis of detailed examination of the human locus, proposed that an initial duplication gave the ancestors of *GH*-like and *CSH*-like genes. A further duplication gave four genes, two *GH*-like and two *CSH*-like and a third duplication gave rise to the third *CSH*-like gene (*hCSH-A*). The organization of the orangutan and gorilla *GH* gene clusters described here accords well with the first two steps in this scheme - the situation in orangutan agrees with that proposed by Chen et al. (1989) after the second round of duplication, with apparent correspondence between *oGH-N*:*hGH-N*, *oCSH-L*:*hCSH-L*, *oGH-V*:*hGH-V* and *oCSH-B*:*hCSH-B*. However, the origin of the additional *CSH* gene(s) in human, chimpanzee and gorilla is less clear.

Phylogenetic analysis of the *GH*/*CSH*-like genes in these four species is shown in Fig. 6A, based on an alignment (~2600 nt) of gene sequences from ~800 nt upstream of the initiating ATG to ~220 nt downstream of the stop codon (beyond these limits similarity between *GH* and *CSH* sequences is lost). The tree in Fig. 6A was constructed using parsimony in PAUP*; likelihood or distance methods gave similar results. The tree shows clustering of *GH-N* and *GH-V* genes, confirming orthology of these genes as expected. However, the clustering of *oCSH-L* and *oCSH-B* that is seen is not as expected. Clustering of *hCSH-L*, *gCSH-A1* and *cCSH-A1* confirms the expected orthology of these genes. However, inclusion of the other *CSH*s in a clade distinct from these suggests that the 'extra' *CSH* genes following *CSH-L*/*A1* in human, chimpanzee and gorilla arose from *CSH-B*, rather than the expected tandem duplication of *CSH-L*/*A1*.

Apparent relationships within the phylogenetic tree may be confused by gene conversion within the gene cluster. Thus the close relationship between *hCSH-A* and *hCSH-B* in Fig. 6A probably

reflects the large conversion between these two genes (see above). In order to minimize this influence analysis of a longer alignment of *CSH* genes (~16090 nt) was performed, using sequences extending ~4100 nt upstream from the initiating ATG and ~10400 nt downstream of the stop codon. The derived phylogenetic tree is shown in Fig. 6B. Clustering of *oCSH-L* and *oCSH-B* is confirmed. Clustering of the other *CSH*s into three groups (*hCSH-L*, *cCSH-A1*, *gCSH-A1*; *hCSH-A*, *cCSH-A2*, *gCSH-A2, gCSH-A3*; *hCSH-B*, *cCSH-B*, *gCSH-B*) is now much as expected, but again the *CSH-A/A2* group clusters with the *CSH-B* group, rather than the adjacent *CSH-L/A1* group.

The relationship between *oCSH-L* and *hCSH-L* is not clear. Each is positioned immediately after *GH-N,* and the gene mutation (..GT.. -> ..AT..) at the start of intron 2 (5' splice donor site) that interferes with normal splicing is the same in human and orangutan, suggesting that the two genes are orthologous. However, the phylogenetic analysis (Fig. 6) does not support such orthology. Either the apparently identical mutation at the splice site arose independently in orangutan and human, or the phylogenetic relationships seen in Fig. 6 are confused by gene conversion. In the latter case, the splice-site mutation in *CSH-L* occurred before separation of lineages leading to orangutan and human, and restoration of this site allowing normal splicing in *cCSH-A1* and *gCSH-A1* was due to reversal of the point mutation or gene conversion.

Evolutionary rate variation

It was shown previously (Perez-Maya et al. 2012) that episodes of rapid, potentially adaptive, evolution occurred prior to divergence of the *CSH* genes and of the *GH-V* genes. Addition of genes for orangutan and gorilla allows these episodes to be defined more fully. An alignment of

CDS sequences was analysed using the codeml method in paml (Yang 2007) with all branches unconstrained and a defined tree; this allows evolution of synonymous (dS) and nonsynonymous (dN) sites in a coding sequence to be distinguished. The latter affect protein sequence and are normally subject to stringent purifying selection, while the former have no effect on protein sequence and are mostly adaptively neutral and little affected by purifying selection (Yang and Bielawski 2000). As a consequence, dN/dS is usually << 1.0. dN/dS is expected to approach 1.0 if a gene loses function and purifying selection no longer applies. In the case of adaptive evolution (positive selection) dN/dS is expected to increase; if dN/dS significantly exceeds 1.0 positive selection is clearly established.

Data were analysed using three codeml models and a defined phylogenetic tree. Detailed results are given in Supplementary Table 1. Model 1 allows a different dN/dS on each branch, and derivation of phylogenetic trees based on dN and dS values (Fig. 7). Rapid evolution on the branch preceding divergence of CSHs is apparent in the tree based on dN values. dN/dS for this branch was 9.27 suggesting adaptive evolution (significance tested later). Significant variation of dN/dS values was demonstrated from the log Likelihood (lnL) test (Supplementary Table. 1). Model 2 allows 2 dN/dS values and was used to test the significance of the rapid evolution on the branch to CSHs. Using this model dN/dS on this branch was 13.98, compared with 0.287 on all other branches. With the appropriate NULL model, in which dN/dS on the branch to CSHs was fixed at 1.0, the lnL test, showed that dN/dS on the branch to CSHs was significantly greater than 1.0 ($2 \times \Delta\text{lnL} = 5.78$; $P<0.05$), indicating positive selection on this branch. Model A of the branch-site method (Zhang et al. 2005), which allows identification of residues probably subject to positive selection, gave dN/dS 30.2, again significantly >1.0 ($2 \times \Delta\text{lnL}= 6.54$; $P<0.05$). The

branch-site method defined 29 residues, substitution of which probably involved positive selection (Supplementary Table 1); these agree closely with those appearing to change on the basis of sequence comparisons (Fig. 2).

Rapid evolution prior to divergence of *GH-V* genes is also apparent, with elevated dN/dS levels; but this accelerated evolution continues after divergence of *oGH-V*. dN/dS levels here do not exceed 1.0, so adaptive evolution is not proven, though it is not excluded (and seems likely given the different biological roles and properties of GH-N and GH-V).

Conclusions

The complex *GH* gene locus in hominids arose as a consequence of gene duplications occurring during the course of primate evolution. As a result of this recent appearance and its complex nature the locus has continued to show evolutionary change during hominid evolution, although some general features were established after the initial two gene duplications. Thus, all hominids appear to have a functioning gene for pituitary GH (*GH-N*), placental GH (*GH-V*) and at least one functioning gene for CSH/PL. In human and orangutan there is only one functional CSH protein - in human because *CSH-A* and *CSH-B* encode the same mature protein and although *CSH-L* may encode a CSH-like protein, this would not be secreted and *CSH-L* may be a pseudogene, but in orangutan because only *CSH-B* encodes a CSH while *CSH-L* is a pseudogene. However, chimpanzee and gorilla respectively have genes that can potentially

produce three and four distinct CSHs. Whether there are functional differences between these different CSHs in gorilla and chimpanzee remains to be established.

Acknowledgements This research was possible thanks to grants contributed by CONACYT (Q-43987). We thank Sergio Lozano-Rodriguez, MD, for his critical reading of the manuscript, and Irám Sánchez for his help in isolating *GH* loci-containing BACs from CHORI genome projects BAC resource, accessed thanks to the generosity of Peter de Jong, Ph D.

References

Adkins RM, Nekrutenko A, Li WH (2001) Bushbaby growth hormone is much more similar to nonprimate growth hormones than to rhesus monkey and human growth hormones. Mol Biol Evol 18: 55-60

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410

Barrera-Saldaña HA, Seeburg PH, Saunders GF (1983) Two structurally different genes produce the same secreted human placental lactogen hormone. J Biol Chem 258: 3787–3793

Barrera-Saldaña HA (1998) Growth hormone and placental lactogen: biology, medicine and biotechnology. Gene 211: 11-8

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, Albert FW, Zeller U, Khaitovich P, Grützner F, Bergmann S, Nielsen R, Pääbo S, Kaessmann H (2011) The evolution of gene expression levels in mammalian organs. Nature 478: 343-348.

Chen EY, Liao YC, Smith DH, Barrera-Saldaña HA, Gelinas RE, Seeburg PH (1989) The

human growth hormone locus: nucleotide sequence, biology, and evolution. Genomics 4: 479–497

De Vos AM, Ultsch M, Kossiakoff AA (1992) Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. Science 255: 306-312

Eberhardt NL, Jiang SW, Shepard AR, Arnold AM, Trujillo MA (1996) Hormonal and cell-specific regulation of the human growth hormone and chorionic somatomammotropin genes. Prog Nucleic Acid Res Mol Biol 54: 127–163

Elefant F, Su Y, Liebhaber SA, Cooke NE (2000) Patterns of histone acetylation suggest dual pathways for gene activation by a bifunctional locus control region. EMBO J 19: 6814–6822

Emera D, Wagner GP (2012) Transposable element recruitments in the mammalian placenta: impacts and mechanisms. Briefings in Functional Genomics 11: 267-276

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783-791

Glass CK, Franco R, Weinberger C, Albert VR, Evans RM, Rosenfeld MG (1987) A c-*erb-A* binding-site in rat growth hormone gene mediates trans-activation by thyroid-hormone. Nature 329: 738-741

González Alvarez R, Revol de Mendoza A, Esquivel Escobedo D, Corrales Félix G, Rodríguez Sánchez I, González V, Dávila G, Cao Q, de Jong P, Fu YX, Barrera Saldaña HA (2006) Growth hormone locus expands and diverges after the separation of New and Old World Monkeys. Gene 380: 38-45

Harper ME, Barrera-Saldaña HA, Saunders GF (1982) Chromosomal localization of the human placental lactogen-growth hormone gene cluster to 17q22-24. Am J Hum Genet 34: 227-

234

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73: 237–244

Ho Y, Liebhaber SA, Cooke NE (2004) Activation of the human GH gene cluster: roles for targeted chromatin modification. Trends Endocrinol Metab 15: 40-45

Jacquemin P, Alsat E, Oury C, Belayew A, Muller M, Evain-Brion D, Martial JA (1996) The enhancers of the human placental lactogen B, A, and L genes: progressive activation during *in vitro* trophoblast differentiation and importance of the DF-3 element in determining their respective activities. DNA Cell Biol 15: 845-54

Jacquemin P, Oury C, Peers B, Morin A, Belayew A, Martial JA (1994) Characterization of a single strong tissue-specific enhancer downstream from the three human genes encoding placental lactogen. Mol Cell Biol 14: 93–103

Jiang SW, Shepard AR, Eberhardt NL (1995) An initiator element is required for maximal human chorionic somatomammotropin gene promoter and enhancer function. J Biol Chem 270: 3683-3692

Kent WJ (2002) BLAT – The BLAST-Like Alignment tool. Genome Res 12: 656–664

Leidig F, Shepard AR, Zhang W, Stelter A, Cattini PA, Baxter JD, Eberhardt NL (1992) Thyroid hormone responsiveness in human growth hormone-related genes. Possible correlation with receptor-induced DNA conformational changes. J Biol Chem 267: 913-921

Lemaigre FP, Courtois SJ, Lafontaine DA, Rousseau GG (1989) Evidence that the upstream stimulatory factor and the Sp1 transcription factor bind *in vitro* to the promoter of the human-growth-hormone gene. Eur J Biochem 181: 555-561

Locke DP and 100 other authors (2011) Comparative and demographic analysis of orang-utan

genomes. Nature 469:529-533

Lynch VJ, Leclerc RD, May G, Wagner GP (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nature Genetics 11: 1154-1159

Lynch VJ, Nnamani MC, Kapusta A, Brayer K, Plaza, SL, Mazur EC, Emera D, Sheikh SZ, Grutzner, F, Bauersachs S, Graf A, Young SL, Lieb JD, DeMayo FJ, Feschotte C, Wagner GP (2015) Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. Cell Reports 10: 1-11

Lytras A, Surabhi RM, Zhang JF, Jin Y, Cattini PA (1996) 'Repair' of the chorionic somatomammotropin-A 'enhancer' region reveals a novel functional element in the chorionic somatomammotropin-B enhancer. Mol Cell Endocrinol 119: 1-10

Misra-Press A, Cooke NE, Liebhaber SA (1994) Complex alternative splicing partially inactivates the human chorionic somatomammotropin-like (*hCS-L*) gene. J Biol Chem 269: 23220–23229

Nachtigal MW, Nickel BE, Cattini PA (1993) Pituitary-specific repression of placental members of the human growth hormone gene family. A possible mechanism for locus regulation. J Biol Chem 268: 8473-8479

Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grützner F, Kaessmann H (2014) The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature 505:635-640.

Norquay LD, Yang X, Jin Y, Detillieux KA, Cattini PA (2006) Hepatocyte nuclear factor-3α binding at P sequences of the human growth hormone locus is associated with pituitary repressor function. Mol Endocrinol 20: 598-607

Norquay LD, Yang X, Sheppard P, Gregoire S, Dodd JG, Reith W, Cattini PA (2003) RFX1 and NF-1 associate with P sequences of the human growth hormone locus in pituitary chromatin. Mol Endocrinol 17: 1027-1038

Pérez-Ibave DC, Rodríguez-Sánchez IP, Garza-Rodríguez ML, Barrera-Saldaña HA (2014) Extrapituitary growth hormone synthesis in humans. Growth Hormone & IGF Research 24: 47-53.

Pérez-Maya AA, Rodríguez-Sánchez IP, de Jong P, Wallis M, Barrera-Saldaña HA (2012) The chimpanzee GH locus: composition, organization, and evolution. Mamm Genome 23: 387-98.

Pipes L, Li S, Bozinoski M, Palermo R, Peng X, Blood P, Kelly S, Weiss JM, Thierry-Mieg J, Thierry-Mieg D, Zumbo P, Chen R, Schroth GP, Mason CE, Katze MG (2013) The non-human primate reference transcriptome resource (NHPRTR) for comparative functional genomics. Nucl Acids Res 41: D906-914.

Reséndez-Pérez D, Ramirez-Solís R, Varela-Echavarría A, Martínez-Rodríguez HG, Barrera-Saldaña HA (1990) Coding potential of transfected human placental lactogen genes. Nucl Acids Res 18: 4665-4670

Rodríguez-Sánchez IP, Tejero ME, Cole SA, Comuzzie AG, Nathanielsz PW, Wallis M, Barrera-Saldaña HA (2010) Growth hormone-related genes from baboon (*Papio hamadryas*): Characterization, placental expression and evolutionary aspects. Gene 450: 1-7

Sawyer SA (1999) GENECONV: a computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University, St. Louis, MO; available at: http://www.math.wustl.edu/_sawyer

Scally A and 71 other authors (2012) Insights into hominid evolution from the gorilla genome

sequence. Nature 483: 169-175

Slater EP, Rabenau O, Karin M, Baxter JD, Beato M (1985) Glucocorticoid receptor binding and activation of a heterologous promoter by dexamethasone by the first intron of the human growth hormone gene. Mol Cell Biol 5: 2984-2992

Smit AFA, Hubley R, Green P (1996-2010) RepeatMasker Open-3.0. Available at http://www.repeatmasker.org

Su Y, Liebhaber SA, Cooke NE (2000) The human growth hormone gene cluster locus control region supports position-independent pituitary- and placenta-specific expression in the transgenic mouse. J Biol Chem 275: 7902-7909

Swofford DL (1998) PAUP*. Phylogenetic analysis using parsimony (* and other methods), ver. 4. Sinauer, Sunderland

Tamura K, Peterson D, Peterson N, Stecher G, Nei M. Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739

Walker WH, Fitzpatrick SL, Barrera-Saldaña HA, Reséndez-Perez D, Saunders GF (1991) The human placental lactogen genes: Structure, function, evolution and transcriptional regulation. Endocrine Rev. 12: 316-328

Wallis M (2008) Mammalian genome projects reveal new growth hormone (GH) sequences. Characterization of the GH-encoding genes of armadillo (*Dasypus novemcinctus*), hedgehog (*Erinaceus europaeus*), bat (*Myotis lucifugus*), hyrax (*Procavia capensis*), shrew (*Sorex araneus*), ground squirrel (*Spermophilus tridecemlineatus*), elephant (*Loxodonta africana*), cat (*Felis catus*) and opossum (*Monodelphis domestica*). Gen Comp Endocrinol 155: 271–279

Wallis OC, Wallis M (2002) Characterization of the GH gene cluster in a new-world monkey, the marmoset (*Callithrix jacchus*). J Mol Endocrinol 29: 89-97

Wallis OC, Wallis M (2006) Evolution of growth hormone in primates: the GH gene clusters of the New World Monkeys marmoset (*Callithrix jacchus*) and white-fronted capuchin (*Cebus albifrons*). J Mol Evol 63: 591–601

Wallis OC, Zhang YP, Wallis M (2001) Molecular evolution of GH in primates: characterization of the GH genes from slow loris and marmoset defines an episode of rapid evolutionary change. J Mol Endocrinol 26: 249–258

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586-1591.

Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15: 496-503

Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol 22: 2472-2479

TABLES

Table 1. Gene conversion between *GH/CSH* genes - based on short gene sequences.

GENECONV was run with mismatches allowed (setting g1).

Sequence pairs showing gene conversion	nt Positions within alignment*	Length of gene conversion detected	Location within genes	Sim *p* Value**	BC KA *p* Value**
Gorilla					
CSH-A1:CSH-A2	1330-2338	1009	exon2-exon5	0.0004	0.0025
CSH-A1:CSH-A3	1330-2320	991	exon2-exon5	0.0127	0.054
GH-N:CSH-A1	347-861	515	5' upstream	0.0137	0.055
CSH-A1:CSH-B	1330-2338	1009	exon2-exon5	0.0153	0.061
Orangutan					
GH-N:GH-V	1511-1741	231	intron2-intron3	0.024	0.058
GH-N:GH-V	1295-1431	137	exon2-intron2	0.024	0.058

***** The alignments analysed were 2661 nt long (gorilla; ATG start codon at 925-927 and TAG stop codon at 2432-2334) and 2704 nt long (orangutan; ATG start codon at 958-960 and TAG stop codon at 2483-2485) (see Supplementary Figs 3 and 4).

**Sim p values are based on permutations; BC KA p values are Bonferroni Karlin-Altschul p values (Sawyer 1999). All *p* values are corrected for multiple hypothesis testing.

Table 2. Gene conversion between *CSH* sequences - based on long gene sequences

GENECONV was run with mismatches allowed (setting g2)

Sequence pairs showing gene conversion	nt Positions within alignment*	Length of gene conversion detected	Location relative to genes	Sim *p* Value**	BC KA *p* Value**
Human					
hCSH-A:hCSH-B	3383-5159	1777	5' upstream-intron4	<0.0001	1.6×10^{-19}
hCSH-A:hCSH-B	14689-15585	897	3' downstream	<0.0001	3.6×10^{-12}
Chimpanzee					
cCSH-A2:cCSH-B	4125-5127	913	intron 1-intron4	<0.0001	7.4×10^{-5}
cCSH-A1:cCSH-A2	5308-5718	411	exon5-3' downstream	0.0012	0.0014
cCSH-A2:cCSH-B	15037-15462	426	3' downstream	0.0021	0.0031
cCSH-A2:cCSH-B	2454-2908	455	5' upstream	0.0032	0.0042
cCSH-A2:cCSH-B	5365-6093	729	exon5-3' downstream	0.0145	0.0198
cCSH-A2:cCSH-B	3304-3595	292	5' upstream	0.0197	0.0271
cCSH-A1:cCSH-B	11052-11364	313	3' downstream	0.0316	0.04289
Gorilla					
gCSH-A3:gCSH-B	2686-4901	2216	5' upstream-intron3	<0.0001	3.8×10^{-18}
gCSH-A2:gCSH-B	3426-5596	2171	5' upstream-3' downstream	<0.0001	3.5×10^{-13}
gCSH-A2:gCSH-A3	2357-7794	5438	5' upstream-3' downstream	<0.0001	3.6×10^{-11}
gCSH-A2:gCSH-A3	1-742	742	5' upstream	<0.0001	1.1×10^{-7}
gCSH-A1:gCSH-A2	15399-16093	695	3' downstream	<0.0001	0.0001
gCSH-A1:gCSH-A3	15692-16093	402	3' downstream	0.0052	0.0072
gCSH-A1:gCSH-A2	10987-11364	378	3' downstream	0.0110	0.0146
Orangutan					
oCSH-L:oCSH-B	8009-8992	984	3' downstream	0.0057	0.0078

*The alignment analysed was 16093 nt long, ATG start codon at 4095-4097, TAG stop codon at 5587-5589 (see Supplementary Fig. 5).

**Sim p values are based on permutations; BC KA p values are Bonferroni Karlin-Altschul p values (Sawyer 1999). All *p* values are corrected for multiple hypothesis testing.

Table 3. Expression of *GH*-like genes in gorilla transcriptomes

			Hits in Blast search with *hGH* CDS		
Project	Experiment	Tissue	*gGH-N*	*gGH-V*	*gCSH-A1-3, gCSH-B*
SRP007412	SRX081954	testis	1	2	24*
SRP017959	SRX217686	testis	8	3	57**
SRP021223	SRX270635, SRX270637	mixed***	34	20	3

* Of the 24 hits 1 could be specifically assigned to *gCSH-A3*, 4 to *gCSH-B*, and 15 to *gCSH-A2*/ *gCSH-A3/gCSH-B*.

** Of the 57 hits 1 could be specifically assigned to *gCSH-A1*, 4 to *gCSH-B* and 46 to *gCSH-A2*/ *gCSH-A3/gCSH-B*

*** mixed tissues including ovary, but not pituitary, placenta or testis; specific *CSH*s not identifiable

Figure legends

Fig. 1. The genomic organization of the *GH* locus in human, chimpanzee, gorilla and orangutan. A. The human *GH* locus (based on Chen et al. 1989). B. The chimpanzee *GH* locus (based on Pérez-Maya et al. 2012). C. The gorilla *GH* locus based on the present work (Supplementary Fig. 1). D. The orangutan *GH* locus also based on the present work (Supplementary Fig. 2). The boxes in gray indicate the genes of *GH*-like sequences. *CD79B* and *TCAM* (TCAM1 homolog pseudogene) are the genes immediately flanking the *GH* locus in human, chimpanzee, gorilla and orangutan assemblies. In human, the *GH* locus is found on chromosome 17 (17q24.2) and is shown here in reverse orientation. Arrows indicate direction of transcription for orangutan; the same direction applies for other species

Fig. 2. Alignment of amino acid sequences derived from *GH*-like genes. hGH-N was used as a reference; identities to this in other sequences are shown as . and deletions as -. * is a stop codon. The signal peptide sequence is highlighted in gray. The sequences given for hCSH-L and oCSH-L are conceptual translations, assuming splicing equivalent to that seen in the other genes; in practice alternatively spliced forms predominate for *hCSH-L* and probably *oCSH-L*, and it is not certain whether these are translated. The sequence of a prosimian GH (*Nycticebus pygmaeus*, nGH) is shown for comparison. Numbers (top line) correspond to the residues underlined, and refer to the hGH-N sequence.

Fig. 3. Alignment of proximal promoters. The *hGH-N* gene was used as a reference; identities to this are shown as . and deletions as -. Regulatory elements that have been identified in the *hGH-*

N gene are shown in boxes. Numbers (bottom line) correspond to the residues underlined (top line), and are relative to the initiating ATG as 1-3.

Fig. 4. Alignment of the placental enhancers. The sequences of the putative placental enhancers of the gorilla and orangutan loci were aligned with their counterparts in human and chimpanzee. The *hCSH-B* gene was used as a reference; identities to this in other sequences are shown as . and deletions as -. The rectangles show the four DF domains of the placental enhancer. Nucleotides that when mutated cause a reduction in activity of the enhancer are shown as * (Lytras et al. 1996) and $ (Jacquemin et al. 1996). The heavy line at the bottom shows the position of the 5' part of the *MER5B* transposable element (see text). The 3' sequence of this element (66 nucleotides) extends beyond the end of DF4 and is not shown.

Fig. 5. Alignment of the P-elements. The sequences of the putative pituitary inhibitors of the gorilla and orangutan loci were aligned with their counterparts in human and chimpanzee. The *hCSH-A* gene was used as a reference; identities to this in other sequences are shown as . and deletions as -. The rectangles show three regions of protein binding (PSE-A, PSE-B and PSE-C) of the P-element (Norquay et al. 2006). Putative binding sites for HNF-3 and C/EBP are underlined. Shown is the 263 nucleotide sequence identified by Nachtigal et al. (1993) as carrying most of the repressor activity within the larger P-element.

Fig. 6. Phylogenetic trees for *GH*-like genes. A: Tree derived using an alignment of sequences from ~800 nt upstream of exon 1 to ~220 nt downstream from the stop codon. B: Tree derived for just *CSH* genes using an alignment of sequences from ~4100 nt upstream of exon 1 to

~10400 nt downstream from the stop codon. Trees were derived using parsimony in PAUP*. Numbers at nodes are percentage of 1000 bootstrap replicates (Felsenstein 1985) supporting that clade.

Fig.7. Phylogenetic trees for coding sequences for *GH*-like genes. dN and dS values were derived using codeml and a defined tree. Note relatively constant evolutionary rate for dS, but considerable rate variation for dN. The branch to *CSHs* following initial duplication of the *GH* gene is shown as a thick line, and illustrates the accelerated change in dN compared with dS characteristic of adaptive evolution. *GH* (*nGH*) of the prosimian *Nycticebus pygmaeus* (Wallis et al. 2001) was included as an outgroup.

Fig. 1.

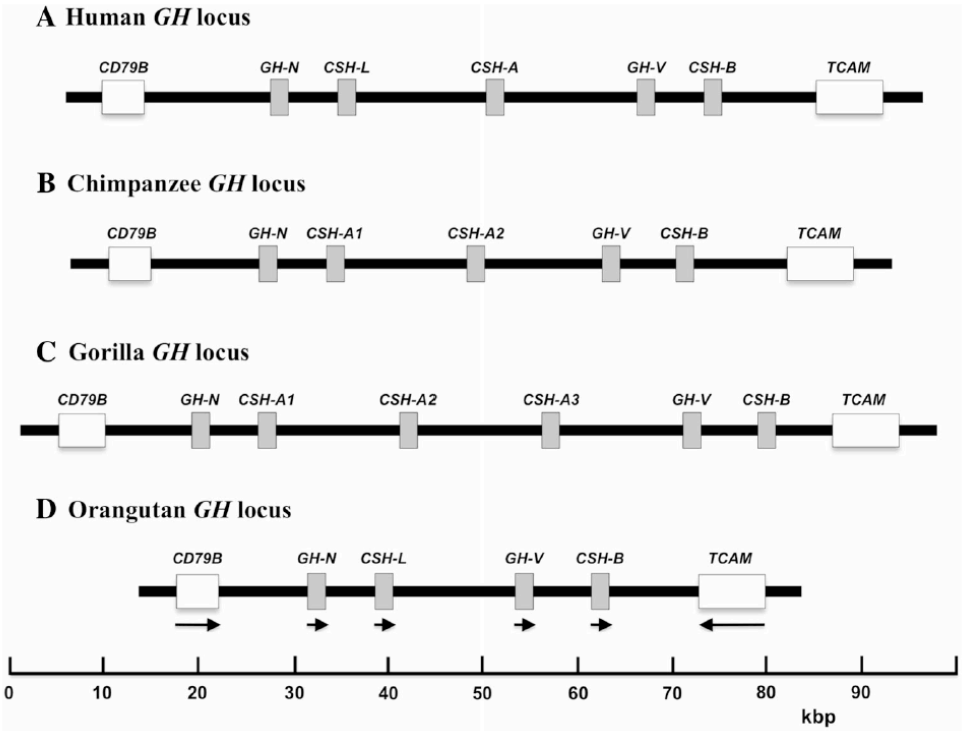

A Human GH locus
CD79B
GH-N
CSH-L
CSH-A
GH-V
CSH-B
TCAM
B Chimpanzee GH locus
CD79B
GH-N
CSH-A1
CSH-A2
GH-V
CSH-B
TCAM
C Gorilla GH locus
CD79B
GH-N
CSH-A1
CSH-A2
CSH-A3
GH-V
CSH-B
TCAM
D Orangutan GH locus
CD79B
GH-N
CSH-L
GH-V
CSH-B
TCAM
0
10
20
30
40
50
60
70
80
90
kbp

Fig. 2.

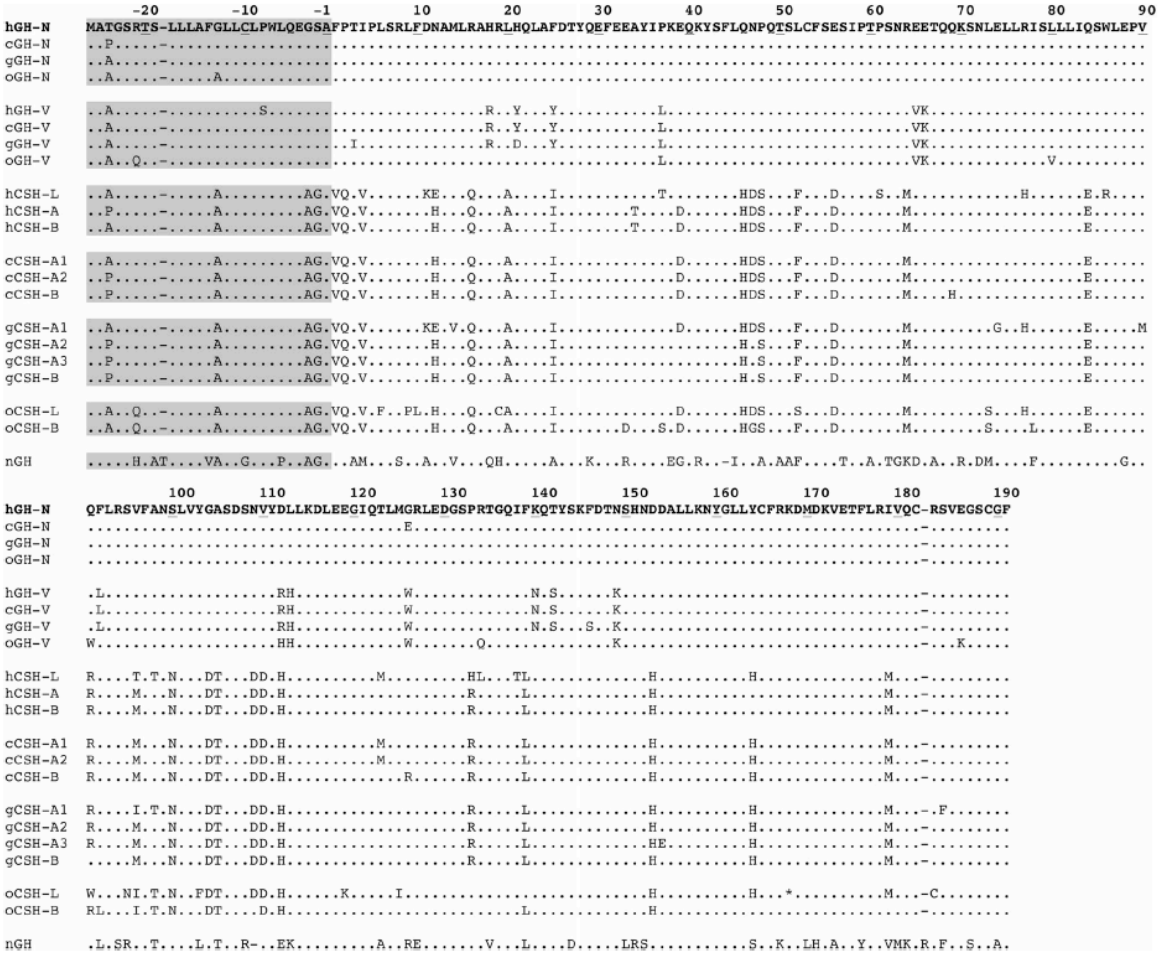

-20 -10 -1 10 20 30 40 50 60 70 80 90
hGH-N MATGSRTS-LLLAFGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLCFSESIPTPSNREETQQKSNLELLRISLLLIQSWLEPV
cGH-N
gGH-N
oGH-N
hGH-V
cGH-V
gGH-V
oGH-V
hCSH-L
hCSH-A
hCSH-B
cCSH-A1
cCSH-A2
cCSH-B
gCSH-A1
gCSH-A2
gCSH-A3
gCSH-B
oCSH-L
oCSH-B
nGH
100 110 120 130 140 150 160 170 180 190
hGH-N QFLRSVFANSLVYGASDSNVYDLLKDLEEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDDALLKNYGLLYCFRKDMDKVETFLRIVQC-RSVEGSCGF
cGH-N
gGH-N
oGH-N
hGH-V
cGH-V
gGH-V
oGH-V
hCSH-L
hCSH-A
hCSH-B
cCSH-A1
cCSH-A2
cCSH-B
gCSH-A1
gCSH-A2
gCSH-A3
gCSH-B
oCSH-L
oCSH-B
nGH

Fig. 3.

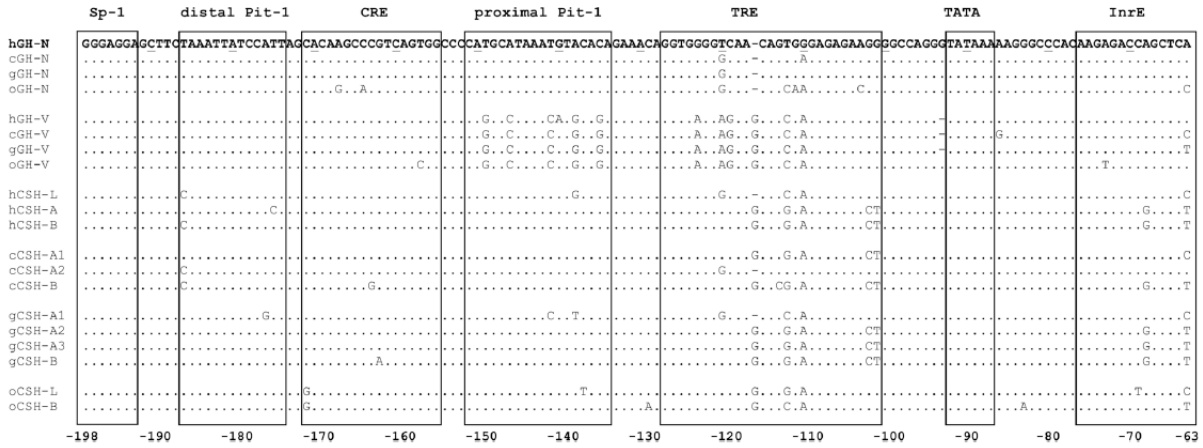

Sp-1
distal Pit-1
CRE
proximal Pit-1
TRE
TATA
InrE
hGH-N GGGAGGAGCTTCTAAATTATCCATTAGCACAAGCCCGTCAGTGGCCCCATGCATAAATGTACACAGAAACAGGTGGGGTCAA-CAGTGGGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCAGCTCA
cGH-N
gGH-N
oGH-N
hGH-V
cGH-V
gGH-V
oGH-V
hCSH-L
hCSH-A
hCSH-B
cCSH-A1
cCSH-A2
cCSH-B
gCSH-A1
gCSH-A2
gCSH-A3
gCSH-B
oCSH-L
oCSH-B
-198 -190 -180 -170 -160 -150 -140 -130 -120 -110 -100 -90 -80 -70 -63

Fig. 4.

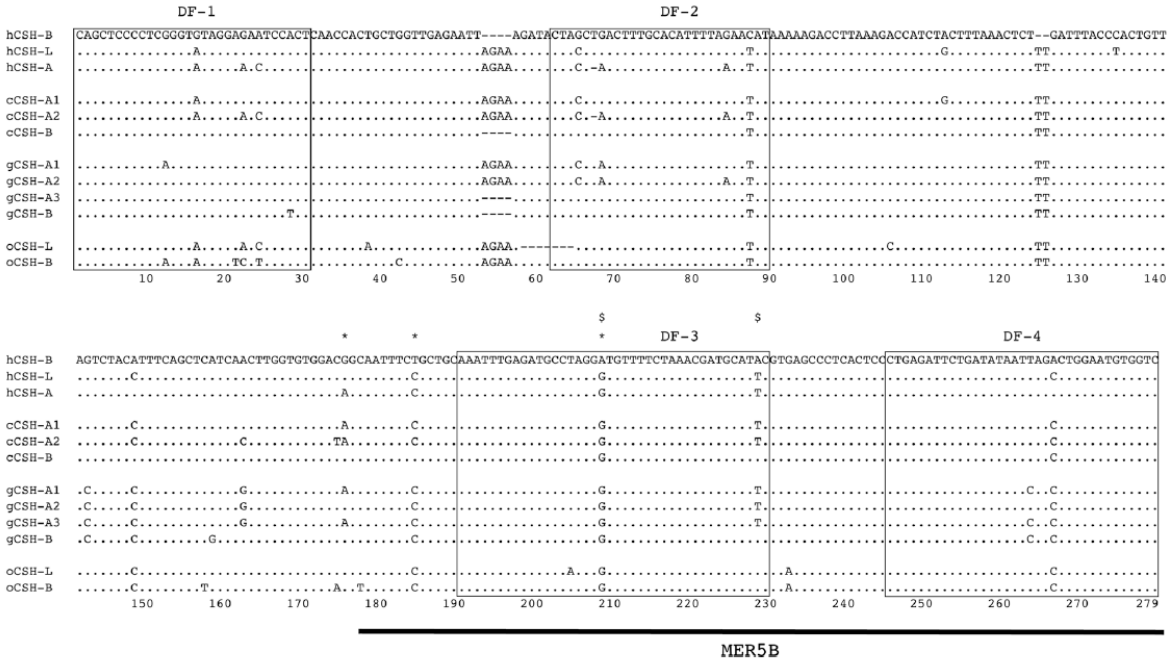

DF-1
DF-2
hCSH-B CAGCTCCCCTCGGGTGTAGGAGAATCCACTCAACCACTGCTGGTTGAGAATT----AGATACTAGCTGACTTTGCACATTTTAGAACATAAAAAGACCTTAAAGACCATCTACTTTAAACTCT--GATTTACCCACTGTT
hCSH-L
hCSH-A
cCSH-A1
cCSH-A2
cCSH-B
gCSH-A1
gCSH-A2
gCSH-A3
gCSH-B
oCSH-L
oCSH-B
10 20 30 40 50 60 70 80 90 100 110 120 130 140
DF-3
DF-4
hCSH-B AGTCTACATTTCAGCTCATCAACTTGGTGTGGACGGCAATTTCTGCTGCAAATTTGAGATGCCTAGGATGTTTTCTAAACGATGCATACGTGAGCCCTCACTCCCTGAGATTCTGATATAATTAGACTGGAATGTGGTC
hCSH-L
hCSH-A
cCSH-A1
cCSH-A2
cCSH-B
gCSH-A1
gCSH-A2
gCSH-A3
gCSH-B
oCSH-L
oCSH-B
150 160 170 180 190 200 210 220 230 240 250 260 270 279
MER5B

Fig. 5.

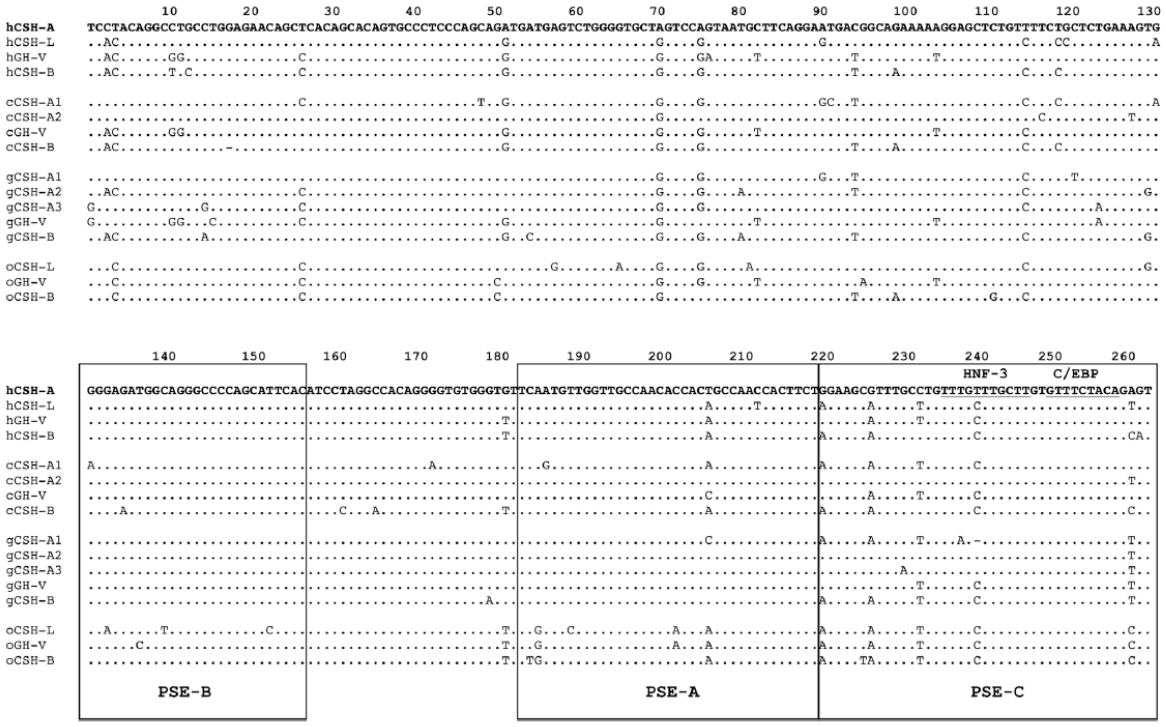

hCSH-A TCCTACAGGCCTGCCTGGAGAACAGCTCACAGCACAGTGCCCTCCCAGCAGATGATGAGTCTGGGGTGCTAGTCCAGTAATGCTTCAGGAATGACGGCAGAAAAAGGAGCTCTGTTTTCTGCTCTGAAAGTG
hCSH-L
hGH-V
hCSH-B
cCSH-A1
cCSH-A2
cGH-V
cCSH-B
gCSH-A1
gCSH-A2
gCSH-A3
gGH-V
gCSH-B
oCSH-L
oGH-V
oCSH-B
HNF-3
C/EBP
PSE-B
PSE-A
PSE-C

Fig. 6.

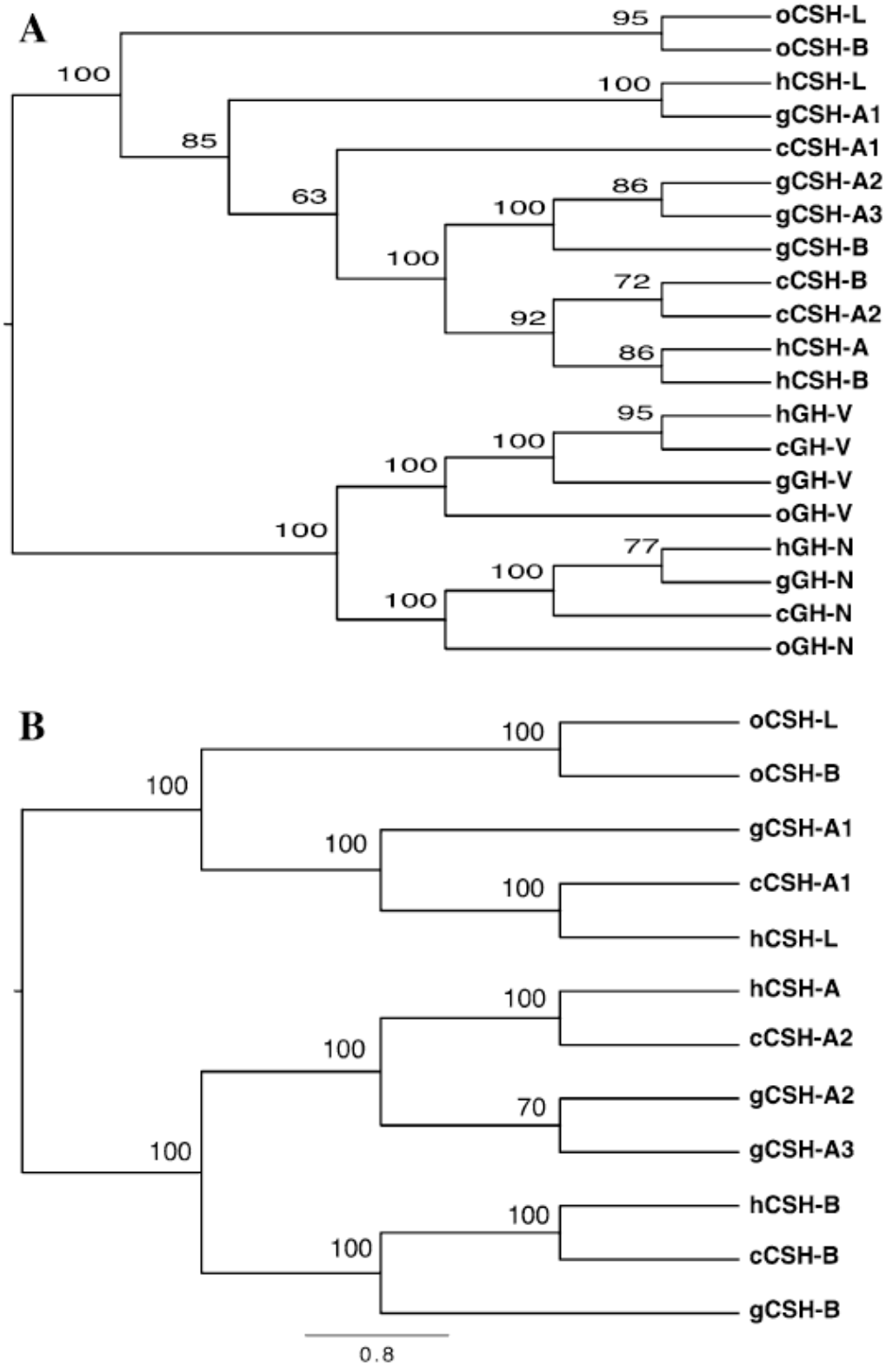
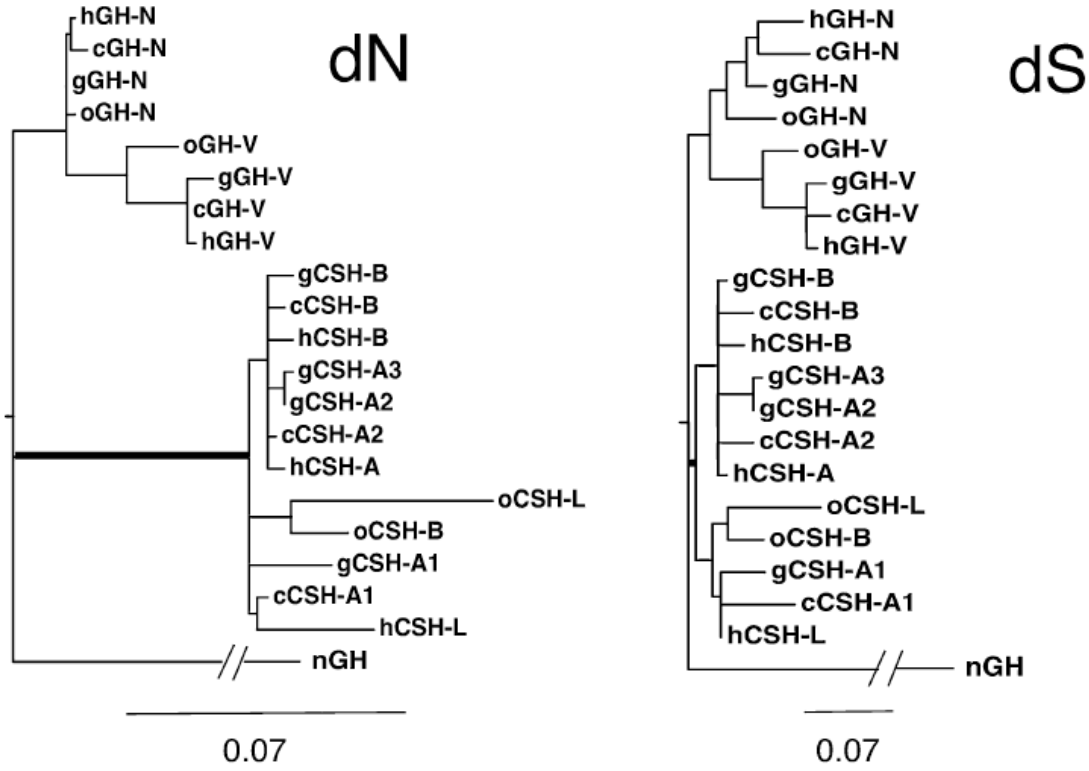

A
95
oCSH-L
oCSH-B
100
100
hCSH-L
gCSH-A1
85
cCSH-A1
86
63
gCSH-A2
100
gCSH-A3
100
gCSH-B
72
cCSH-B
92
cCSH-A2
86
hCSH-A
hCSH-B
95
hGH-V
100
cGH-V
100
gGH-V
oGH-V
100
77
hGH-N
100
gGH-N
100
cGH-N
oGH-N
B
100
oCSH-L
oCSH-B
100
gCSH-A1
100
cCSH-A1
hCSH-L
100
hCSH-A
100
cCSH-A2
70
gCSH-A2
100
gCSH-A3
100
hCSH-B
100
cCSH-B
gCSH-B
0.8

Fig. 7.


dN
hGH-N
cGH-N
gGH-N
oGH-N
oGH-V
gGH-V
cGH-V
hGH-V
gCSH-B
cCSH-B
hCSH-B
gCSH-A3
gCSH-A2
cCSH-A2
hCSH-A
oCSH-L
oCSH-B
gCSH-A1
cCSH-A1
hCSH-L
nGH
0.07
dS
hGH-N
cGH-N
gGH-N
oGH-N
oGH-V
gGH-V
cGH-V
hGH-V
gCSH-B
cCSH-B
hCSH-B
gCSH-A3
gCSH-A2
cCSH-A2
hCSH-A
oCSH-L
oCSH-B
gCSH-A1
cCSH-A1
hCSH-L
nGH
0.07